



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-227228

FY06 LDRD Final Report "Development of Computational Techniques For Decoding The Language of Genomes"

I.V. Ovcharenko

January 12, 2007

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

FY06 LDRD Final Report
“Development of Computational Techniques For
Decoding The Language of Genomes”
LDRD Project Tracking Code: 06-ERD-004
Ivan Ovcharenko, Principal Investigator

Abstract

This project was aimed at developing computational methods and tools to decipher the universal language of gene regulation encoded in genomes. It was proposed to identify and decipher the code and to functionally annotate genomic elements that orchestrate the temporal and spatial dynamics of gene expression in living cells.

Introduction/Background

Currently, no means exists to annotate pathways of protein-DNA interactions, predict biological response to environmental conditions, or explain dynamics of developmental processes on a genome scale. The proposed study will provide the means to tackle these problems computationally and will produce results with immediate practical use in a variety of experimental studies across the Laboratory.

Research Activities and the Summary of Results:

06-ERD-004 was funded by the Computation and Biosciences Directorates in May of 2006 and ended on September 30, 2006. Despite a significantly reduced budget, we were able to accomplish 3 out of 4 FY06 aims. Also, we published 3 manuscripts that were mainly based on results of this ER research (one of these manuscripts will be featured on the cover of the Genome Research journal in February of 2007). In addition to that, we prepared and submitted 2 NIH proposals seeking external funding to continue the research initiated by this LDRD. Specifically, we:

1. combined experimental gene expression profiling and heuristic sequence analysis to define signatures of regulatory elements in complex genomes;
2. developed a computational method for de novo identification of functional elements in genomes;
3. demonstrated that this unique method performs exceptionally well in blind detection of experimentally characterized mouse enhancer elements.

Submitted Proposals:

- NIH R01 grant proposal “Deciphering Principles of Regulatory Genomics”; pending (scored 174 (16.5 top percentile) on Oct. 5, 2006). PI: Ivan Ovcharenko, Co-PI: Gabriela Loots (CMLS).
- NIH R21 grant proposal “Computational tools for decoding vertebrate genome regulation”; pending (to be scored in Jan. 2007). PI: Ivan Ovcharenko

Published manuscripts:

- L.A. Pennacchio, G.G. Loots, M.A. Nobrega, and **I. Ovcharenko**, Predicting Tissue-Specific Enhancers in the Human Genome, Genome Research, 17(1), 2007 (advance online publication on January 8, 2007) (Note: this article will be featured on the cover of the journal.)

- G.G. Loots and **I. Ovcharenko**, ECRbase: Database of Evolutionary Conserved Regions, Promoters, and Transcription Factor Binding Sites in Vertebrate Genomes., *Bioinformatics*, 23(1), 122-124 (2007).
- G.G. Loots, P.S.G. Chain, S. Mabery, A. Rasley, E. Garcia, and **I. Ovcharenko**, Array2BIO: from microarray expression data to functional annotation of co-regulated genes, *BMC Bioinformatics*, 7, 307-311 (2006)